Thoughts on Data Formats for JetScape

Chimera Framework
Generalized Nuclear Data (GND)
HEP Database

Ron Soltz
Nuclear & Particle Physics, LLNL

April 12, 2017

LLNL-PRES-XXXXXX
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

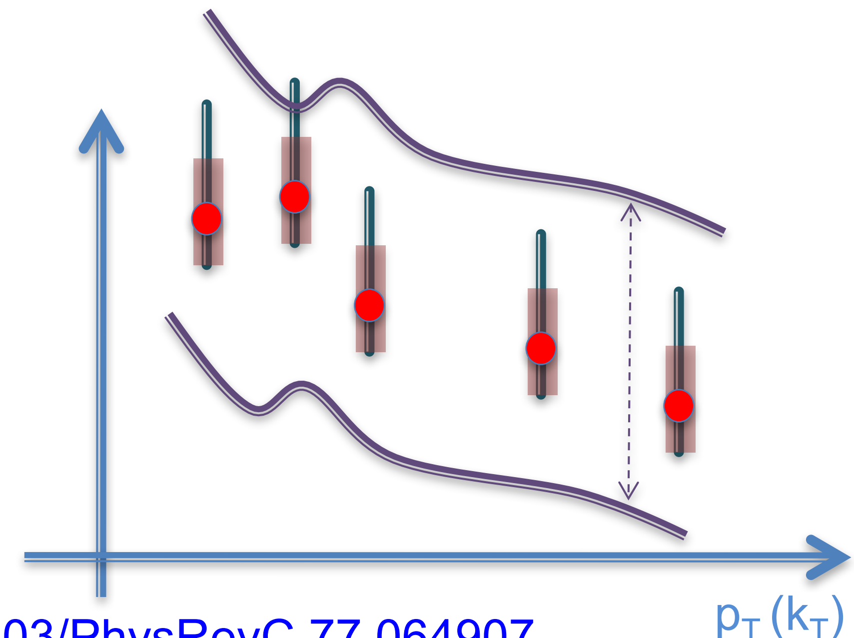
Lawrence Livermore National Laboratory

CHIMERA Framework

- Format developed to handle simultaneous model-to-data chi-squared evaluations for PHENIX and STAR spectra, flow, HBT
- Documented in https://dx.doi.org/10.1103/PhysRevC.87.044901
- Calculate χ^2 from data, account for type A & B errors

- A type: uncorrelated (σ_i)
- B type: correlated frac.(σ_b)
- C type: normalization (σ_c)
- D type: correlated tilt (not considered)

$$\tilde{\chi}^2(\epsilon_b, \epsilon_c, p) = \left[\left(\sum_{i=1}^{n} \frac{(y_i + \epsilon_b \sigma_{b_i} + \epsilon_c y_i \sigma_c - \mu_i(p))^2}{\tilde{\sigma}_i^2}\right) + \epsilon_b^2 + \epsilon_c^2\right]$$



Error definitions based on https://dx.doi.org/10.1103/PhysRevC.77.064907

Chimera Data Format ASCII Example

- PhysRevLett.91.182301_fig3.txt *file name references figure*
  - V2 200 AuAu -211 -321 10 PHENIX 10.1103/PhysRevLett.91.182301 Fig_6 kFullTriangleUp SumPiKwithPiSysErr *1st line specifies measurement*
  - [20,40] Np 140.4 (4.9,4.9) Nc 296.8 (31.1,31.1) *2nd line specifies centrality*
  - 0.210 [0.00,0.25] 0.0225837 (0.2090470E-02,0.2090470E-02) (15%,15%)
  - 0.364 [0.25,0.50] 0.0453414 (0.1044580E-02,0.1044580E-02) (8%,8%)
  - 0.608 [0.50,0.75] 0.0766495 (0.1398750E-02,0.1398750E-02) (6%,6%)
  - 0.859 [0.75,1.00] 0.1077820 (0.1954680E-02,0.1954680E-02) (5%,5%)
  - 1.110 [1.00,1.25] 0.1324270 (0.2710530E-02,0.2710530E-02) (5%,5%)
  - 1.360 [1.25,1.50] 0.1520260 (0.3759720E-02,0.3759720E-02) (5%,5%)
  - 1.691 [1.50,2.00] 0.1742140 (0.4507390E-02,0.4507390E-02) (5%,5%)
  - 2.176 [2.00,2.50] 0.1883490 (0.8535240E-02,0.8535240E-02) (8%,8%)
  - 2.672 [2.50,3.00] 0.1932730 (0.1522460E-01,0.1522460E-01) (11%,11%)
  - 3.274 [3.00,4.00] 0.1668760 (0.2649510E-01,0.2649510E-01) (11%,11%)

p_T [lo,hi], v2 (stat-lo, stat-hi) [sys-lo, sys-hi]

Spectra Example

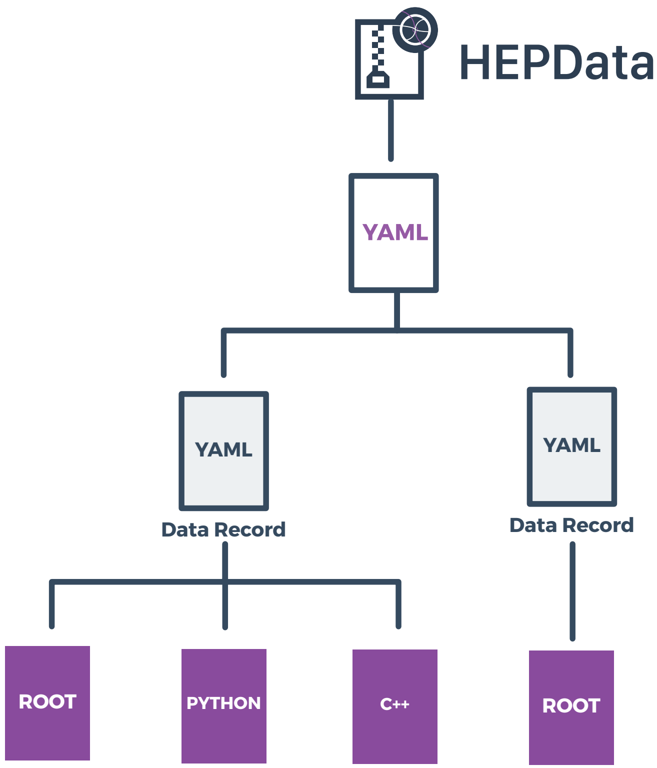
- PhysRevLett.92.112301_fig1.txt
  - Spectra 200 AuAu 211 0 11 STAR 10.1103/PhysRevLett.92.112301 fig_1a kOpenCircle not_FD_corr
  - [10,20] Np 234.6 (9.3,8.3) Nc 591.3 (59.9,51.9)
  - 0.2259 [0.2009,0.2509] 242 (-7.3,7.3) (5%,5%)
  - 0.2752 [0.2502,0.3002] 182.1 (-3.7,3.7) (5%,5%)
  - 0.3249 [0.2999,0.3499] 142 (-2.8,2.8) (5%,5%)
  - 0.3755 [0.3505,0.4005] 111.2 (-1.1,1.1) (5%,5%)
  - 0.4253 [0.4003,0.4503] 87.09 (-0.88,0.88) (5%,5%)
  - 0.4755 [0.4505,0.5005] 68.45 (-0.7,0.7) (5%,5%)
  - 0.5253 [0.5003,0.5503] 53.85 (-0.55,0.55) (5%,5%)
  - 0.5749 [0.5499,0.5999] 42.52 (-0.73,0.73) (5%,5%)
  - 0.6252 [0.6002,0.6502] 33.41 (-0.91,0.91) (5%,5%)
  - 0.6753 [0.6503,0.7003] 26.55 (-0.72,0.72) (5%,5%)
  - 0.7253 [0.7003,0.7503] 21.28 (-0.79,0.79) (5%,5%)

Durham HEPData (now hosted by CERN)

- https://hepdata.net/
- Has many attractive features
  - Searching and plotting
  - Accommodates multiple (point-to-point?) systematic errors
  - Links to papers and analysis code
  - Integrates with ROOT, Python, Mathematica, …

HepData Figure from https://indico.cern.ch/event/512652/timetable

HEPData



submission.yaml
links the submission together by detailing the data files to be loaded, their name and description, and their assocated analysis files and code.

data records
YAML (or JSON) representation of the underlying data files including value errors in a verbose format.

external data files & links
analysis files, code, links to code repositories, etc.

Generalized Nuclear Data Structure

- Developed for low energy nuclear reactor design and nuclear stockpile communities
- GND developed at LLNL with ASC/ARRA funding, web site hosted by BNL Nuclear Data Group
  - https://ndclx4.bnl.gov/gf/project/gnd/
- Documented in Nuclear Data Sheets
  - https://doi.org/10.1016/j.nds.2012.11.008 - initial development
  - http://dx.doi.org/10.1016/j.nds.2014.12.007 - covariance matrices
- Nuclear Data Structure can be implemented in XML, HDF5, …
- Supports both raw and *evaluated* data, including 1-3 dimensional function containers

GND Diagrams for data and covariances

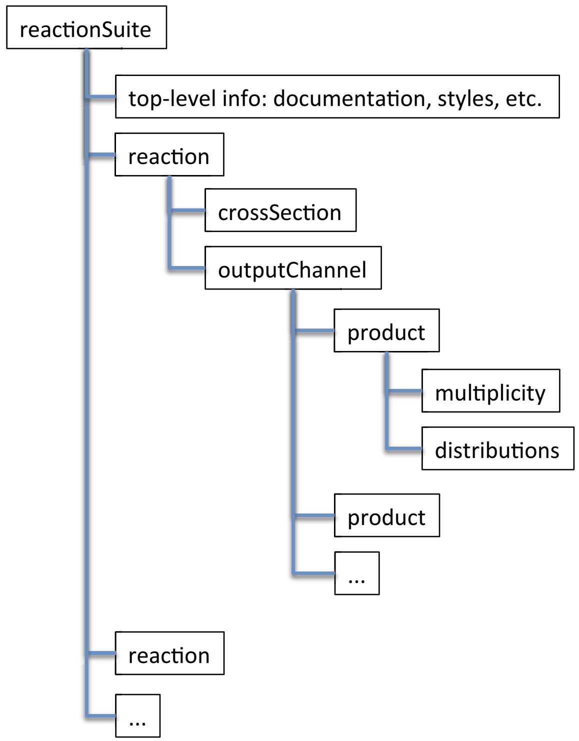


FIG. 1. A simplified overview of how data are organized inside the GND 'reactionSuite'. The nested hierarchy includes multiple 'reaction' elements, each of which contains a cross section and a list of outgoing products along with multiplicities, distributions and possible other data.

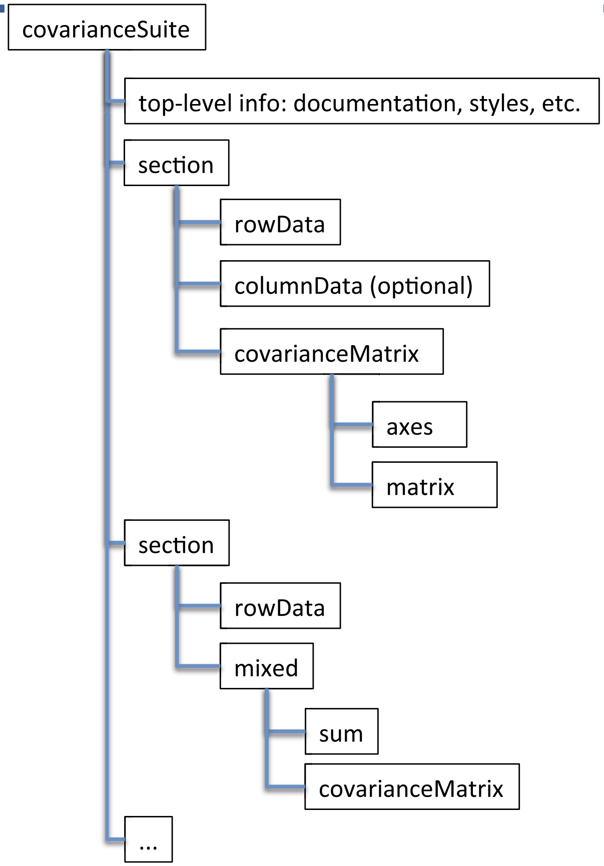


FIG. 3. Inside the covarianceSuite, each rectangular block of the full covariance matrix is stored in a separate section. Sections are identified by their 'rowData' and 'columnData' (which contain links to the cross sections or other data that are being covaried in this section). The section also contains one or more covariance matrices, that may be stored in various forms.

GND Covariance Matrix Block Diagram

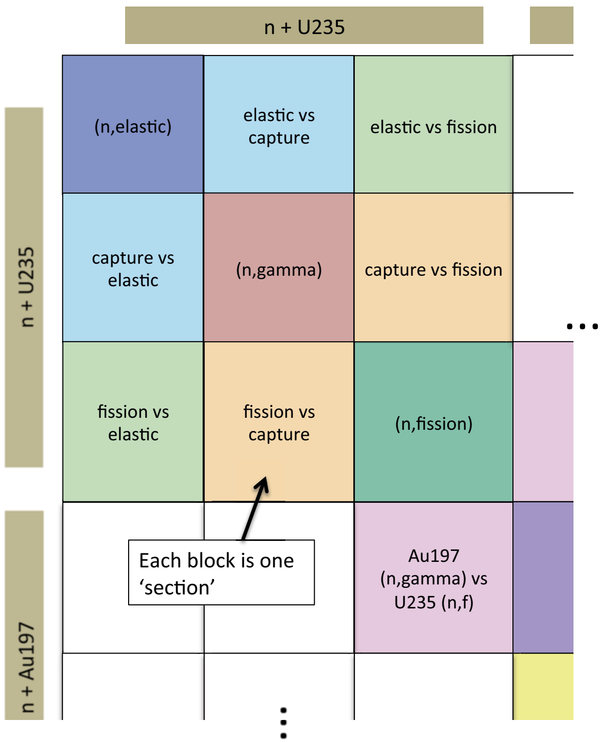


FIG. 2. (Color online) Covariances can be broken up into rectangular sections, each of which contains the covariance between one or two different quantities. These rectangular sections are the primary organization for the GND covarianceSuite.

GND Covariance Matrix Description

- The actual covariance matrix data can be stored inside the section in several different forms, including:
- A single matrix, including a list of energies defining the group boundaries along each axis as well as the list of matrix elements. The matrix includes a flag telling whether its elements correspond to absolute covariances, relative covariances or correlation coef-ficients (an example of this kind of section is shown in Fig. 4),
- a matrix where the axes correspond to model parameters rather than a range of incident energies (currently used mainly to store resonance parameter covariances),
- special storage schemes that use less space to store symmetric, diagonal or sparse covariance matrices,
- a ‘mixed’ covariance, stored as two or more covariance matrices (representing for example statistical and systematic uncertainty contributions, or covariances in two disjoint energy regions) that must be summed together to produce the full matrix, or
- a weighted sum of covariance matrices from other sections (like the NC-type section described in chapter 33 of Ref. [1].)

GND Covariance Matrix XML Example

```xml
<covarianceSuite projectile="n" target="Mn55" evaluation="ENDF/B-8.0" version="GND 1.7">
 - <section label="2n + Mn54 + photon">
     <rowData ENDF_MFMT="33,16" xlink:href="/reactionSuite/reactions/reaction[@label='2n + Mn54 + photon']/crossSection/XYs1d[@label='eval']"/>
   - <covarianceMatrix label="eval" type="relative">
     - <gridded2d>
       - <axes>
         - <grid index="2" label="row_energy_bounds" unit="eV" style="boundaries">
           - <values length="10">
               8971600 1.1295e7 1.4219e7 1.7901e7 2.2536e7 2.8371e7 3.5717e7 4.4965e7 5.6607e7 6e7
             </values>
           </grid>
         - <grid index="1" label="column_energy_bounds" unit="eV" style="link">
             <link xlink:href="../../grid[@index='2']/values"/>
           </grid>
           <axis index="0" label="matrix_elements" unit=""/>
         </axes>
       - <array shape="9,9" symmetry="lower">
         - <values length="45">
             0.9009 0 0.0298027 0 0.0274589 0.0275798 0 0.0255486 0.0253597 0.0272668 0 0.022247 0.0225891 0.0260009 0.0307842 0 0.0290332 0.0278611 0.0249922 0.026839 0.0844295 0 0.0477075
             0.0440729 0.0335427 0.0320437 0.170327 0.393075 0 0.0418178 0.0385906 0.0302851 0.0343094 0.168644 0.398216 0.458782 0 0.0206091 0.0189457 0.0141119 0.0169872 0.0928324
             0.223123 0.275132 0.174932
           </values>
         </array>
       </gridded2d>
     </covarianceMatrix>
   </section>
   ...
</covarianceSuite>
```

JetScape Data Structure – desired features

- Interface to community standard (HEPdata)
- Allow for correlated errors (matrices)
- Comparison to multiple measurements with different systematic errors (can be same exp.) via one of two paths:
  1. Simultaneous comparison to all results
  2. Single comparison to evaluated results – requires extra step
- Accommodate results from experiments, models, and evaluations
- Open Questions (i.e. tasks)
  - Does HEPdata have all data and information we need for JetScape ?
  - Would GND provide a helpful starting point ?